

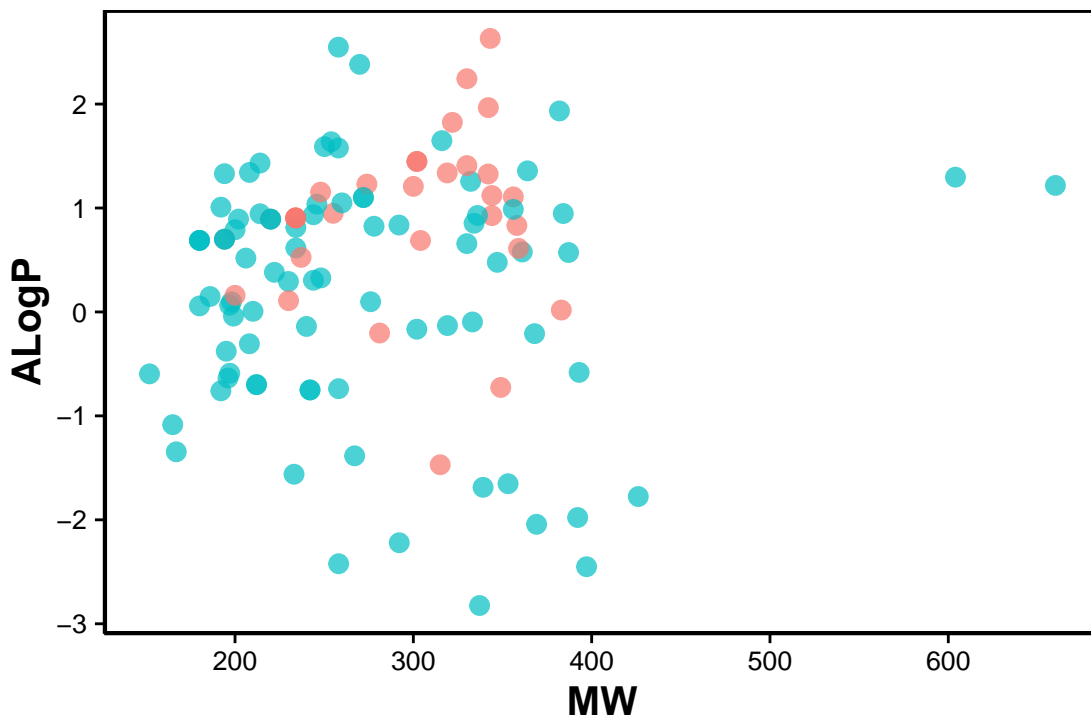
Chemical Space

Chuleeporn Phanus-umporn, Saw Simeon, Watshara Shoombuatong and Chanin Nantasenamat

July 13, 2559 BE

Chemical space of anti-sickling

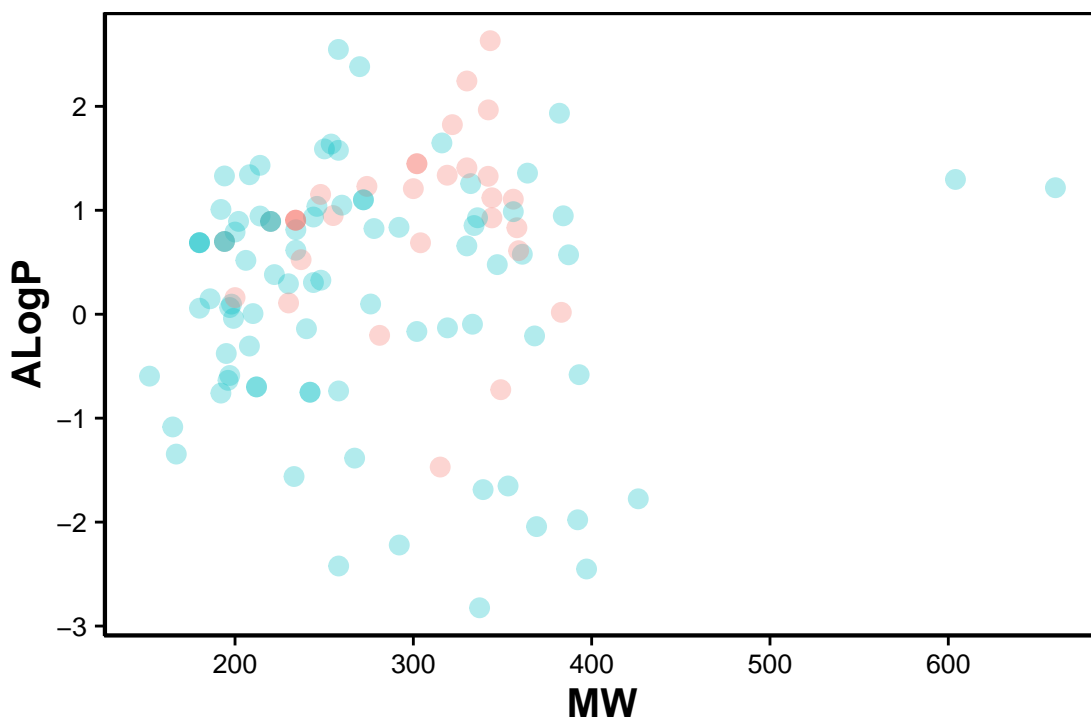
```
library(cowplot)
data <- read.csv("Linpiski_Descriptors.csv")
descriptors_name <- c("ALogP", "nHBAcc", "nHBDdon", "MW")
descriptors <- data[, descriptors_name]
Activity <- data$Activity
data_2 <- cbind(Activity, descriptors)
data_plot <- data.frame(data_2)
p <- ggplot(data_plot, aes(MW, ALogP))
p <- p + geom_point(aes(colour = factor(Activity)), size = 3, alpha = 0.7)
p <- p + theme(legend.position = ("none"),
               panel.border = element_rect(linetype = "solid",
                                             colour = "black", fill = NA, size = 1),
               axis.text.x = element_text(colour = "black", size = 10),
               axis.text.y = element_text(colour = "black", size = 10),
               plot.margin = grid::unit(c(1, 1, 1, 1), "cm"),
               axis.title.x = element_text(colour = "black", size = 15, face = "bold"),
               axis.title.y = element_text(colour = "black", size = 15, face = "bold"))
print(p)
```



Visualization of Lipinski's descriptors

Chemical space of AChE inhibitors are shown as active (green), inactive (red) and intermediate (blue)

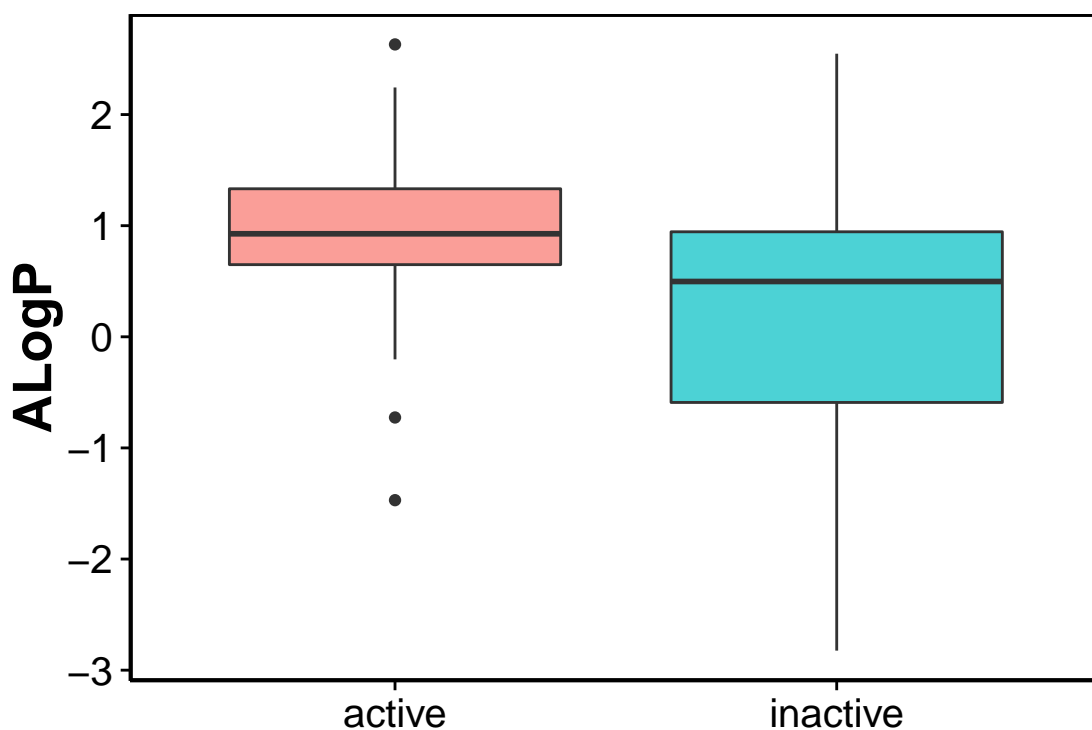
```
library(ggplot2)
library(cowplot)
data_2 <- cbind(Activity, descriptors)
data_2 <- na.omit(data_2)
data_plot <- data.frame(data_2)
p <- ggplot(data_plot, aes(MW, ALogP))
p <- p + geom_point(aes(colour = factor(Activity)), size = 3, alpha = 0.3)
p <- p + theme(legend.position = ("none"),
  panel.border = element_rect(linetype = "solid",
    colour = "black", fill = NA, size = 1),
  axis.text.x = element_text(colour = "black", size = 10),
  axis.text.y = element_text(colour = "black", size = 10),
  plot.margin = unit(c(1, 1, 1, 1), "cm"),
  axis.title.x = element_text(colour = "black", size = 15, face = "bold"),
  axis.title.y = element_text(colour = "black", size = 15, face = "bold")
)
print(p)
```



Boxplot of Lipinski's rule-of-five descriptors (ALogP)

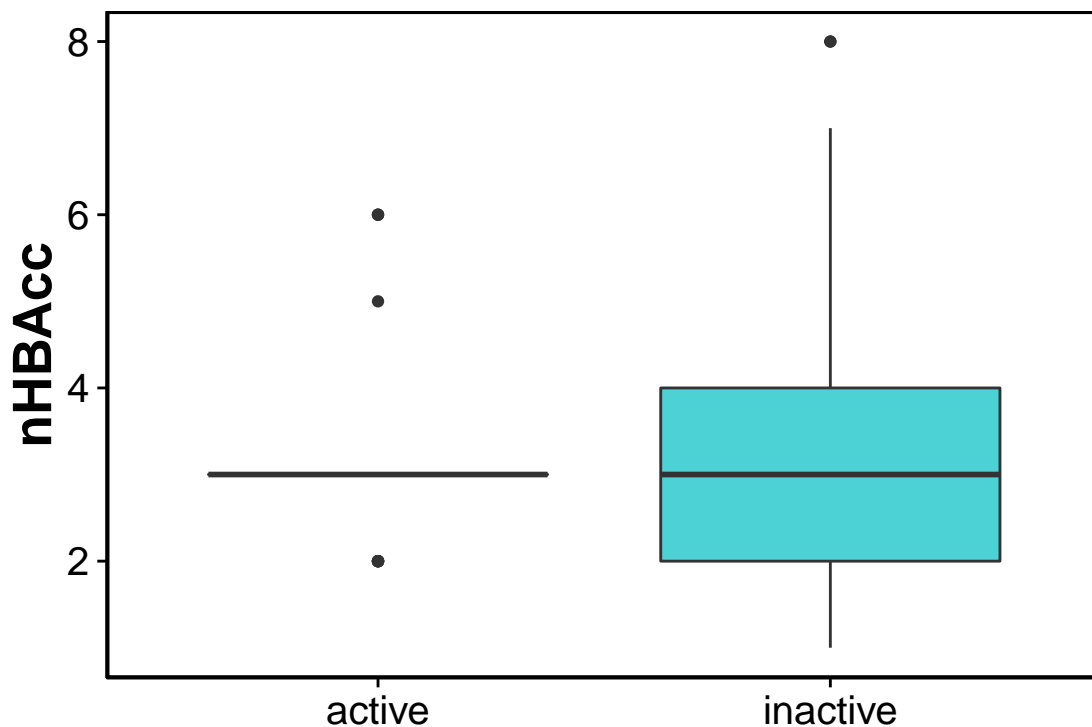
```
p_1 <- ggplot(data_plot, aes(factor(Activity), ALogP))
p_1 <- p_1 + geom_boxplot(aes(fill = factor(Activity)), alpha = 0.7)
p_1 <- p_1 + theme(legend.position = ("none"),
  panel.border = element_rect(linetype = "solid",
    colour = "black", fill = NA, size = 1),
  axis.text.x = element_text(colour = "black", size = 15),
  axis.text.y = element_text(colour = "black", size = 15),
  plot.margin = unit(c(1, 1, 1, 1), "cm"),
  axis.title.y = element_text(size = 20, face = "bold"),
  axis.title.x = element_blank())

print(p_1)
```



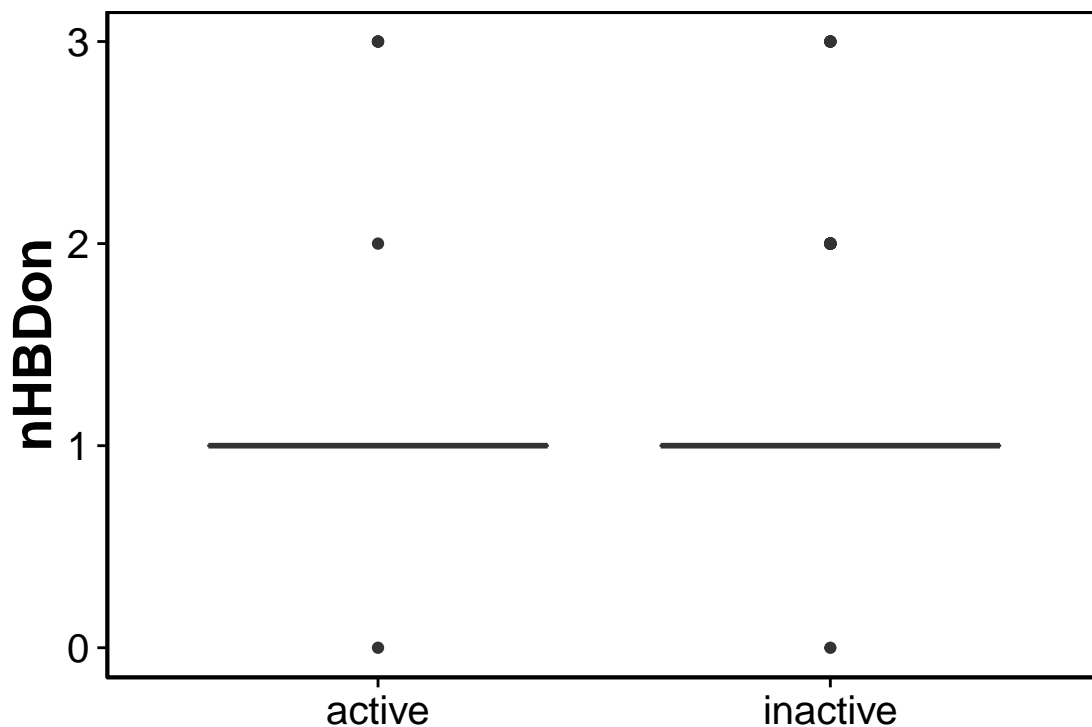
Boxplot of Lipinski's rule-of-five descriptors (nHBAcc)

```
p_2 <- ggplot(data_plot, aes(factor(Activity), nHBAcc))
p_2 <- p_2 + geom_boxplot(aes(fill = factor(Activity)), alpha = 0.7)
p_2 <- p_2 + theme(legend.position = ("none"),
  panel.border = element_rect(linetype = "solid",
    colour = "black", fill = NA, size = 1),
  axis.text.x = element_text(colour = "black", size = 15),
  axis.text.y = element_text(colour = "black", size = 15),
  plot.margin = unit(c(1, 1, 1, 1), "cm"),
  axis.title.y = element_text(size = 20, face = "bold"),
  axis.title.x = element_blank())
print(p_2)
```



Boxplot of Lipinski's rule-of-five descriptors (nHBDon)

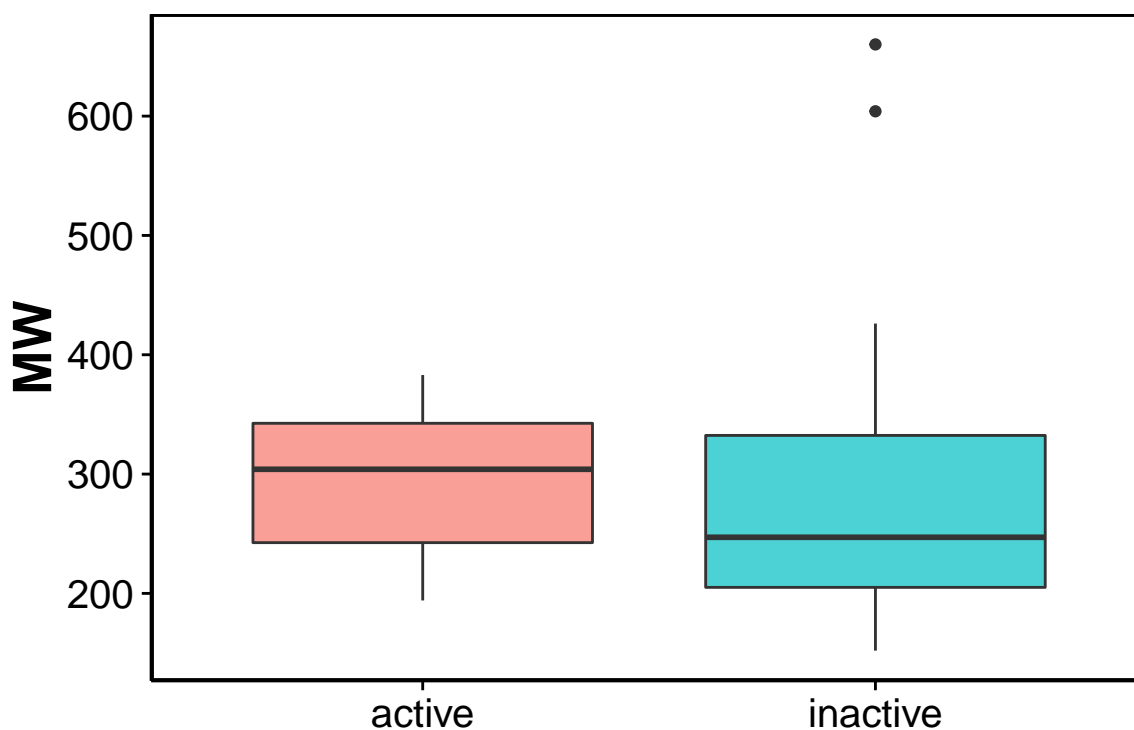
```
p_3 <- ggplot(data_plot, aes(factor(Activity), nHBDon))
p_3 <- p_3 + geom_boxplot(aes(fill = factor(Activity)), alpha = 0.7)
p_3 <- p_3 + theme(legend.position = ("none"),
  panel.border = element_rect(linetype = "solid",
    colour = "black", fill = NA, size = 1),
  axis.text.x = element_text(colour = "black", size = 15),
  axis.text.y = element_text(colour = "black", size = 15),
  plot.margin = unit(c(1, 1, 1, 1), "cm"),
  axis.title.y = element_text(size = 20, face = "bold"),
  axis.title.x = element_blank())
print(p_3)
```



Boxplot of Lipinski's rule-of-five descriptors (MW)

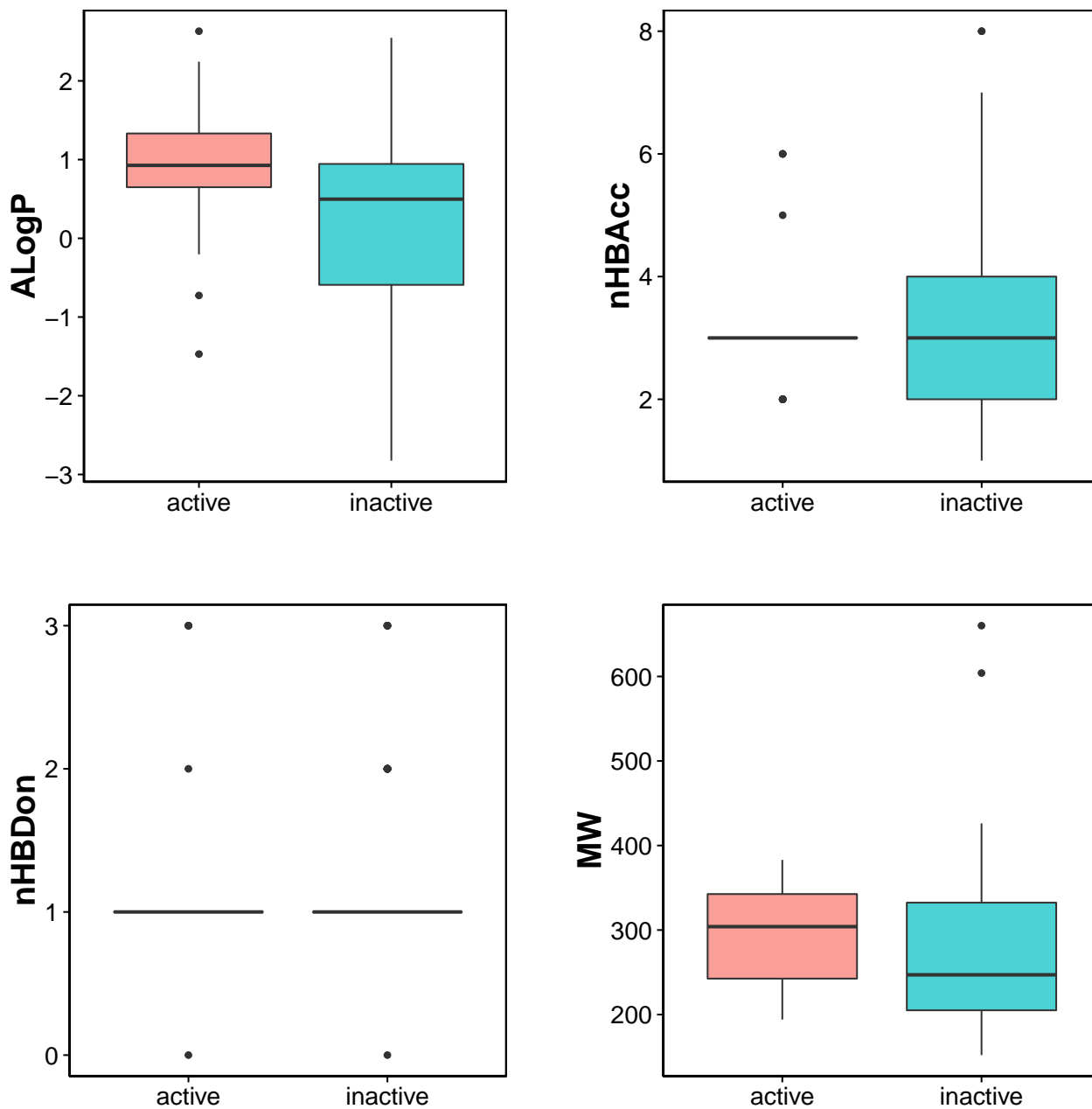
```
p_4 <- ggplot(data_plot, aes(factor(Activity), MW))
p_4 <- p_4 + geom_boxplot(aes(fill = factor(Activity)), alpha = 0.7)
p_4 <- p_4 + theme(legend.position = ("none"),
  panel.border = element_rect(linetype = "solid",
    colour = "black", fill = NA, size = 1),
  axis.text.x = element_text(colour = "black", size = 15),
  axis.text.y = element_text(colour = "black", size = 15),
  plot.margin = unit(c(1, 1, 1, 0.4), "cm"),
  axis.title.y = element_text(size = 20, face = "bold"),
  axis.title.x = element_blank())

print(p_4)
```



Arranging boxplot of Lipinski's rule-of-five descriptors (ALogP, nHBAcc, nHBDOn, MW)

```
library(cowplot)
plot_grid(p_1, p_2, p_3, p_4)
```



EDA of feature important analysis

```
library(caret)
library(reshape2)
```

```

df <- read.csv("data.csv")
names <- c("Activity", "SubFPC171", "SubFPC5", "SubFPC1", "SubFPC298", "SubFPC2", "SubFPC16", "SubFPC17",
          "SubFPC100")
df <- df[, names]
data_melt <- melt(df)
ggplot(aes(y = value, x = variable, fill = Activity), data = data_melt) + geom_boxplot() +
  ylab("Substructure Count") + theme(
    legend.position = "none",
    axis.text.y = element_text(size = 20, colour = "black"),
    axis.text.x = element_text(size = 20, colour = "black"),
    plot.margin = unit(c(1,1, 1, 1),"cm"),
    panel.border = element_rect(linetype = "solid", colour = "black", fill = NA, size = 1),
    axis.title = element_text(size = 25, face = "bold", colour = "black")
  )

```

